AN INVESTIGATION OF THE DYNAMICS OF FORECAST ERRORS AND ITS REPRESENTATION IN THE THORPEX INTERACTIVE GRAND GLOBAL ENSEMBLE (TIGGE)

A Thesis

by

MICHAEL AARON HERRERA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Istvan Szunyogh
Committee Members,	R. Saravanan
	Robert L. Korty
	Robert D. Hetland
Head of Department,	Ping Yang

December 2014

Major Subject: Atmospheric Sciences

ABSTRACT

We employ local linear diagnostics to investigate the efficiency of an ensemble in capturing the space and magnitude of forecast uncertainties. In this study, we use ensemble forecast data from the leading NWP centers included in the THORPEX Interactive Grand Global Ensemble (TIGGE). We find that the operational ensembles included in the TIGGE data set are highly efficient in capturing the space and magnitude of forecast uncertainties past about 72 hour forecast lead time. We also find that our diagnostics are able to provide valuable information on the performance of the ensembles in respect to model setup, such as ensemble generation techniques and model error parameterizations. Lastly we investigate the geographical distribution of ensemble performance for select operational centers. We utilize components of the eddy kinetic energy equation to find how transient processes at the synoptic scale might affect our diagnostics.

DEDICATION

I want to dedicate this thesis to my wife Samantha Phelps-Herrera. Her love and support over the last ten years of our relationship has allowed me to find and pursue my passion for meteorology and the atmosphere. When we started dating in my junior year of high school, I had no idea the direction my life would take but we have risen to face all challenges together and shared in our successes. She is now an amazing high school teacher with an unmatched ability to reach out and connect with at-risk kids and serves as a daily inspiration to be the best person that I can be. It is truly an understatement when I say that I would't be where I am today without her these last ten years.

ACKNOWLEDGEMENTS

I would like to start by thanking my committee chair and advisor, Dr. Istvan Szunyogh for all of his help with not only this research, but helping me to become a better graduate student and researcher. Looking back on the person I was when I entered graduate school, I can not only see how much he has helped me to grow, but also how much room I still have for improvement. Thanks are also due to my other committee members, Dr. R. Saravanan, Dr. Robert Korty, and Dr. Robert Hetland for their valuable time and suggestions on this research project.

I would also like to thank my parents, Alex and Andrea Herrera. My mother has been incredibly supportive of me through all the ups and downs that I've faced in life, from health problems when I was younger to figuring out what I wanted to do in college. My father instilled a sense of curiosity and wonderment of the world since I was little, which is still present to this day. It is with their help and support all my life that I am able to be where I am today. Thank you.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

In the most influential paper ever written on the dynamics of forecast error growth, Lorenz (1969) investigated the role of scale interactions in the error growth process. He argued that forecast errors propagated upscale, leading to a loss of predictability at increasingly larger scales as forecast time increased. Lorenz's famous result was most recently revisited by Tribbia and Baumhefner (2004) and Rotunno and Snyder (2008). Most importantly, the former paper augmented Lorenz's description of the process by the observation that in the extratropics the dominant errors asymptoted to the baroclinically active scales, where they then grew exponentially.

Satterfield and Szunyogh (2010, 2011), hereafter referred to collectively as S10-11, observed a spectral evolution of the errors that was similar to the one reported by Tribbia and Baumhefner (2004). S10-11 also found that a small (40-80-member) ensemble of forecast perturbations was highly efficient, in a linear sense, in capturing the local space of forecast uncertainty for forecast times longer than about 72-h forecast time. Here, the term 'local' refers to the three-dimensional state of the atmosphere in a column of about 1000 km horizontal radius. They also found, however, that on average, the ensemble underestimated the magnitude of the forecast uncertainty. In other words, while the particular combination of model and ensemble generation technique was efficient in capturing the dominant instabilities of the flow, it also underestimated the magnitude of the forecast uncertainty associated with those instabilities.

S10-11 observed the aforementioned behavior of the ensemble in both realistic and perfect model experiments. In the perfect model scenario, the only potential explanation for the underestimation of the forecast uncertainty was the inadequate

representation of the evolution of the initial condition uncertainty. While sub-optimality of the technique for the generation of the initial ensemble perturbations may have played some role, we suspect that the inherent limitations of a small ensemble in representing the nonlinear propagation of the uncertainties between scales played the dominant role in their result. When the model is not perfect, the problem is further complicated by the presence of model errors, which can inject uncertainties at all scales and distort the scale interactions that propagate the uncertainties between the different scales.

A representation of the effect of upscale propagating uncertainties on the magnitude of the uncertainties at the synoptic scales, whose structure is well resolved by the ensemble, require the development of "parameterization" techniques. The techniques for the "parameterization" of the effect of upscale propagating errors are called Stochastic Kinetic Energy Backscatter (SKEB) algorithms (e.g., Shutts 2005, 2013; Berner et al. 2009; Bowler et al. 2009; Charron et al. 2010; Tennant et al. 2011). In addition, the two approaches for the "parameterization" of the complex model errors that inject uncertainty into the forecasts are the Stochastically Perturbed Parameterization Tendencies (SPPT) scheme (Buizza et al. 1999; Palmer et al. 2009) and the family of multi-parameterization (multi-physics) methods (e.g., Berner et al. 2011; Houtekamer 2002).

In the present study, we extend the diagnostic investigations of S10-11 to the operational global ensemble prediction systems of the world by taking advantage of the unique opportunity provided by the THORPEX Interactive Grand Global Ensemble (TIGGE). This data set includes operational global ensemble forecast data from all major operational prediction centers of the world that produce such forecasts. These centers use a variety of carefully tuned schemes to represent the effects of initial condition and model uncertainties.

The outline of the paper is the following. Section 2 provides the necessary background information about the TIGGE data set, while section 3 describes the local diagnostics we use. Section 4 describes the dynamics of the atmosphere for the time period of the study, while section 5 presents the results of the diagnostic calculations that were applied to all ensembles in the TIGGE data set. Section 6 shows further diagnostic results for a select group of ensembles and section 7 offers our conclusions.

2. THE TIGGE DATA SET

We first give a general description of the TIGGE data set, then provide a brief discussion of the techniques that the different centers use for the representation of the effects of initial condition and model uncertainties.

2.1 The data set

The TIGGE data set is a collection of global ensemble forecasts from the major NWP centers of the world. The goal with the creation of the data set was to provide data to support both academic research on predictability and ensemble forecasting, and operational product development at the forecast centers. Ensemble forecasts generated by the NWP centers have been collected in real time at three archive centers (ECMWF, NCAR, and CMA) and made available to the scientific community in an easily accessible uniform format. At the time of writing this paper, ECMWF has made plans to become the main distributor of the data. Users can request data from these centers formatted to their needs, choosing from any of the available forecasts, variables, grids, regions, etc.

The forecast data used in our study are from the following forecast centers:

- European Centre for Medium-Range Weather Forecasts (ECMWF)
- US National Centers for Environmental Prediction (NCEP)
- UK Met Office (UKMO)
- China Meteorological Administration (CMA)
- Japan Meteorological Agency (JMA)
- Korean Meteorological Administration (KMA)

- Meteorological Service of Canada (CMC)
- Météo-France

We do not analyze data from two of the NWP centers, the Australian Bureau of Meteorology (BoM) and the Centro de Previsão de Tempo e Estudos Climáticos (CPTEC), that provide data to theTIGGE data set. These centers were excluded because data was not available from the BoM ensemble for the time period of this study, while CPTEC discovered an error in their ensemble and they are in the process of regenerating the forecast data for the TIGGE data set.

2.2 Initial condition perturbations

A fundamental challenge of ensemble forecasting is that the degrees of freedom of the dynamics of an operational model is orders of magnitude larger than the operationally attainable number of ensemble members. Several methods have been developed to generate initial perturbations that efficiently represent the growing part of the analysis (initial condition) errors. Table B.1 shows how the initial perturbations of the different ensembles included in the paper are generated.

The bred vector method (Toth and Kalnay 1993, 1997) was originally developed and implemented at NCEP. The two centers currently using bred vectors are CMA and KMA. To create the bred vectors, the analysis is randomly perturbed and the full non-linear model is run for a short period (eg. six hours) for both the control (unperturbed) and perturbed analyses. The control forecast is subtracted from the perturbed forecast and then scaled down to the size of the initial perturbation. The cycle is repeated by adding this new scaled perturbation to the new analysis. After several days of "breeding", growing patterns dominate the spatio-temporal evolution of the perturbations. This is a desired feature of the initial perturbations since the uncertainty in the initial state is dominated by fast growing errors.

Another method to generate perturbations, which is used by ECMWF, JMA, and Météo-France is known as (right) singular vectors (Buizza et al. 1993; Molteni and Palmer 1993; Mureau et al. 1993). Right singular vectors are the initial perturbations that grow fastest with respect to a preselected norm and optimization (forecast) time. For the ensembles included in this study, the norm is a quadratic norm with energy dimension (e.g., Buizza et al. 1993) and the optimization time is 48 h forecast hour. While bred vectors are perturbations that grow fastest over an extended period of the immediate past, singular vectors are the perturbations that will grow the fastest for a limited time of the immediate future. Singular vectors are dependent on the choice of norm that is used and for all three ensemble systems that use singular vectors, a total energy norm is used. Using a total energy norm creates an upscale transfer of energy from the initial singular vectors to the final singular vectors at the optimization time. The Météo-France ensemble uses a combination of singular vectors and evolved singular vectors, where the evolved singular vectors are created such that the analysis time, for which the initial perturbations are created, coincides with the end of the optimization period. The evolved singular vectors are hoped to represent analysis uncertainties that were likely to grow in the analysis cycles of the immediate past. ECMWF also used evolved singular vectors in the past, but by the time of the present study they have switched to using an ensemble of data assimilations (EDA) to account for the error growth during the previous data assimilation cycles (Buizza et al. 2008). To create these perturbations, observations are perturbed randomly in accordance with their presumed error statistics in the data assimilation system; each set of perturbed observations is assimilated into a different ensemble member.

The method currently used for the generation of ensemble perturbations at NCEP is similar to the generation of bred vectors, but it uses information from the data assimilation system to determine a spatio-temporally varying rescaling factor. This

method is called Ensemble Transform with Rescaling (ETR) and was developed by Wei et al. (2008). Ensemble perturbations valid for the analysis time are obtained through an ensemble transform of a previous set of forecast perturbations while taking into account the observation statistics and centers the perturbations on the analysis. The end result is similar to bred vectors in which the fastest growing perturbations are sampled while also maintaining the analysis error variance found by the independent data assimilation system.

The UK Met Office uses a local Ensemble Transform Kalman Filter (ETKF) to generate their perturbations, (Bishop et al. 2001; Wang and Bishop 2003; Bowler and Mylne 2009). The largest difference between the ETKF and the ETR methods is that the ETKF produces a full analysis ensemble rather than a spatio-temporally varying rescaling factors. Even though an analysis is also generated by the ETKF, UKMO does not use that analysis. Instead, they center the ETKF analysis perturbations on the operational 4D-Var analysis. Lastly, CMC uses an Ensemble Kalman Filter (EnKF) to generate the analysis ensemble. Unlike at UKMO, their ensemble is centered on the mean analysis produced by the EnKF.

2.3 Model error parameterization techniques

In addition to chaotic model dynamics acting on uncertain initial conditions, model errors also contribute to the forecast error growth. The effects of model errors enter into the forecasts continuously during the entire forecast period. Model errors also contribute to the initial conditions uncertainty through the forecast phase of the analysis cycles.

At the current level of the state of the art in numerical weather prediction, the main sources of model errors are thought to be the parameterization schemes for the sub-grid processes. One technique that is used for the simulation of errors that enter

into the forecasts through the parameterization schemes is Stochastically Perturbed Parameterization Tendencies, SPPT (Buizza et al. 1999; Palmer et al. 2009). This technique perturbs the total contribution of the parameterized process to the tendency of the state variables in the model. Another approach to simulate the uncertainties in the parameterization schemes is to use different parameterization schemes for the same processes, or to use different values of the prescribed parameters in the same scheme. This approach is known as the multi-physics technique (Berner et al. 2011; Houtekamer 2002).

The effects of uncertainties injected at the smallest resolved scales cannot be directly simulated by the models, because the interactions between those scales and the larger scales are distorted by the models; some scale interactions are explicitly eliminated by the truncation strategies, while other are eliminated by dampening the smaller scale motions. Time integration schemes also contribute to the diffusiveness of the models at scales where nature is not diffusive. An approach to make the representation of the effect of the upscale propagating uncertainties by the model more realistic, called Stochastic Energy Backscattering (SKEB), was put forth by Shutts (2005). Since SPPT/multi-physics and SKEB are used to simulate different aspects of model error dynamics, they can (and should) be used in conjunction. This practice is followed at both ECMWF and CMC.

3. LOCAL DIAGNOSTICS

To investigate the ability of an ensemble to capture the magnitude and the spatial structure of the forecast uncertainties, we follow the approach of S10-11, applying linear diagnostics to the ensemble perturbations within local volumes. We choose this approach for two reasons. First, it allows for the introduction of spatially local diagnostic quantities whose definition requires the availability of a vector (linear) space. Second, it eliminates the difficulties we would face with introducing similar diagnostics for the global states due to the rank deficiency of the ensemble-based estimate of the global forecast error covariance matrix.

3.1 Local vectors and their covariance

We define a local state vector $\mathbf{x}_{V_\ell}$ to describe the state in a local volume V_ℓ centered at model grid point ℓ. The components of $\mathbf{x}_{V_\ell}$ are the grid point variables of the model in V_ℓ. We assume the availability of a K-member forecast ensemble and introduce the notation $\mathbf{x}^k_{V_\ell}$, k=1,...,K, for the members of the ensemble of local state vectors. Then, the local ensemble perturbations, $\mathbf{X}^k_{V_\ell}$, k=1,...,K, can be defined as

$$\mathbf{X}^k_{V_\ell} = \mathbf{x}^k_{V_\ell} - \bar{\mathbf{x}}_{V_\ell}, \tag{3.1}$$

where

$$\bar{\mathbf{x}}_{V_\ell} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{x}^k_{V_\ell} \tag{3.2}$$

is the local ensemble mean. In what follows, we treat all local vectors as column vectors.

With the help of the local ensemble perturbations, we can define the ensemble-

based estimate of the local forecast error covariance matrix,

$$\mathbf{P}_{V_\ell} = \frac{1}{K-1} \sum_{k=1}^{K} \mathbf{X}^k_{V_\ell} [\mathbf{X}^k_{V_\ell}]^T. \tag{3.3}$$

The local state vector, perturbations and error covariance matrix can be defined at any forecast lead time t_f, including the analysis time ($t_f = 0$). We treat the spaces of local ensemble perturbations as linear spaces. To be precise, we define the space of local ensemble perturbations at forecast time t_f for the local volume V_ℓ by the range of $\mathbf{P}_{V_\ell}(t_f)$. In this space, which we denote by $\mathbb{S}^K_{V_\ell}(t_f)$, a linear combination of the local ensemble perturbations is also a local perturbation. Because we compute the diagnostics for all local volumes V_ℓ and forecast times t_f, in what follows, we drop the subscripts and the argument from the notation.

3.2 Definition of the diagnostics

We apply diagnostics to the difference

$$\delta \mathbf{x}^t = \mathbf{x}^t - \bar{\mathbf{x}} \tag{3.4}$$

between a proxy $\mathbf{x}^t$ of the local true state and the local ensemble mean, $\bar{\mathbf{x}}$. The difference $\delta \mathbf{x}^t$ is often interpreted as the error in the ensemble mean forecast. This terminology is fully justified when the ensemble mean is used as a deterministic forecast. Here, we strictly view $\delta \mathbf{x}^t$ as a random variable that represents the difference between the state of the atmosphere and the predicted mean of the probability distribution of the local state. We assume, for the time being, that $\bar{\mathbf{x}}$ is an accurate prediction of the mean of the probability distribution of the local state. Under this

assumption, the vector $\delta\mathbf{x}^t$ can be decomposed as

$$\delta\mathbf{x}^t = \delta\mathbf{x}^{t(\|)} + \mathbf{x}^{t(\perp)} \tag{3.5}$$

where $\delta\mathbf{x}^{t(\|)}$ is the component of $\delta\mathbf{x}^t$ that projects onto $\mathbb{S}^K$ and $\mathbf{x}^{t(\perp)}$ is the component that projects onto the null space of $\mathbf{P}$. The set of normalized eigenvectors, $\{\mathbf{u}_k : k = 1, \ldots, K-1\}$, associated with the largest $K-1$ eigenvalues of $\mathbf{P}$, provide a convenient orthonormal basis to compute $\delta\mathbf{x}^{t(\|)}$ by

$$\delta\mathbf{x}^{t(\|)} = \sum_{k=1}^{K-1} \left([\delta\mathbf{x}^t]^T \mathbf{u}_k \right) \mathbf{u}_k. \tag{3.6}$$

We employ three diagnostics based on the concept of the local space of ensemble perturbation. These are the mean-square magnitude

$$TV = E\left(\|\delta\mathbf{x}^t\|^2 \right) = E\left(\left(\delta\mathbf{x}^t\right)^T \delta\mathbf{x}^t \right), \tag{3.7}$$

of $\delta\mathbf{x}$; the quantity

$$TVS = E\left(\|\delta\mathbf{x}^{t(\|)}\|^2 \right) = E\left(\left(\delta\mathbf{x}^{t(\|)}\right)^T \delta\mathbf{x}^{t(\|)} \right), \tag{3.8}$$

which measures the extent to which the linear space spanned by the ensemble perturbations can represent the possible states of the atmosphere; and the local ensemble variance,

$$VS = E\left(\text{trace}(\mathbf{P}) \right). \tag{3.9}$$

In the definition of the diagnostic quantities, $E(\cdot)$ denotes the expected value of the random variables in the argument for the grand ensemble of those random variables for

either all forecasts of the same lead time and all grid points in the verification region, or for all forecasts of the same lead time. In the former case, each diagnostics has a single scalar value for each forecast time, while in the latter case, each diagnostics is a discretzied field of scalar grid point variables for each forecast time.

The diagnostics TV, TVS and VS can be used to define necessary (but not satisfactory) conditions that an accurate ensemble prediction system should satisfy. In particular, for an ensemble that captures all uncertain features in the the forecasts and correctly represents the total magnitude of the uncertainty associated with those features, $TV = TVS = VS$. This is only a necessary condition, because the effects of the errors in the prediction of the probability distribution of the state on the diagnostics at the different times and/or locations can cancel out each other. Yet, the optimal case, in which this condition would be satisfied is rarely observed in practice. An investigation of the ways an ensemble fails to satisfy the condition provides important information about the shortcomings of the ensemble prediction system. In particular,

- if $TV > VS$ ($TV < VS$), the ensemble underestimates (overestimates) the total magnitude of the uncertainty;
- if $TVS > VS$ ($TVS < VS$), the ensemble underestimates (overestimate) the total magnitude of the uncertainty associated with the error patterns that it captures correctly.

We note that an ensemble always satisfies the condition $TV \geq TVS$.

3.3 Experiment design

We compute diagnostics for the forecasts that were started between 1 January, 2012 0000 UTC and 29 February, 2012 1800 UTC. The diagnostics are computed for the

entire forecast range of each ensemble system. Diagnostics that require the estimation of temporal means are computed by taking averages over all forecasts of equal forecast time issued during the two-month period of our investigation. Spatiotemporal means for the NH extratropics are computed by averaging the temporal means over all locations between 30°N and 75°N.

The local volume V_ℓ is defined by the atmospheric column given by 5-by-5 horizontal grid points centered at ℓ, and the levels between the surface and the 200 hPa pressure level. Because the data set has a 2.5×2.5 degree horizontal resolution, the horizontal dimension of a local volume in the midlatitudes is about 1000 km. Components of the local state vector $\mathbf{x}_{V_\ell}$ are defined by the temperature, zonal and meridional wind, and surface pressure grid point variables within the local volume V_ℓ. The different components are scaled such that the Euclidean norm of the local state vectors has dimension of energy (Talagrand 1981; Buizza et al. 1993).

We use ECMWF analyses as the proxy $\mathbf{x}^t$ for the true atmospheric state in most computations. The exceptions are the computations of the diagnostics for the ECMWF and Météo-France ensembles, in which NCEP analyses are used as the proxy for the true state. In the case of the ECMWF ensemble, the NCEP data are used to ensure that the random variables that represent the errors in the verified and the verifying data are statistically independent. In the case of the Météo-France ensemble, the NCEP data are used, because the Météo-France forecasts are for 0600 UTC and 1800 UTC and NCEP is the only center that provides analyses for those times in the TIGGE data set. To test the sensitivity of the results to the choice of the verifying data, diagnostic calculations for additional ensembles were also carried out with using the NCEP analyses as the proxy of the true states.

4. THE ATMOSPHERIC FLOW

Because transient processes at the synoptic scales play a central role in the exponential growth of the dominant error patterns, before we turn our attention to the discussion of the diagnostic results, we provide a description of the dynamics of the transient processes for the time period of our investigation. This description is based on the eddy kinetic energy equation of Orlanski and Katzfey (1991); Orlanski and Chang (1993); Chang (2000). Because not all the variables that are necessary for the computation of the terms of the equation are available in the TIGGE data set, we use data from the ERA Interim reanalysis for the computation of the terms of the equation.

4.1 Definition of the eddy component of the flow

The computation starts with a decomposition of the spatiotemporally evolving atmospheric state variables into a spatially varying time-mean component and a spatiotemporally evolving eddy component. We compute the time-mean for January-February-March, because even though all forecasts investigated here start in January and February, some of them end in March. The time-mean component of the geopotential height field at the 500 Pa pressure level is shown in Fig. A.1: the time-mean flow has a dominantly zonal wavenumber two structure, with negative zonal anomalies in the Pacific and the Atlantic storm track regions and positive zonal anomalies in the exit region of the two storm tracks.

4.2 The eddy kinetic energy equation

The eddy kinetic energy equation in pressure coordinate system is

$$\frac{\partial}{\partial t}\langle K_e\rangle = -\langle\nabla\cdot\mathbf{v}K_e\rangle - \langle\nabla\cdot\mathbf{v}'\phi'\rangle - \langle\omega'\alpha'\rangle - \langle\mathbf{v}'\cdot(\mathbf{v}'_{\mathbf{3}}\cdot\nabla_3)\mathbf{v}_{\mathbf{m}} - \mathbf{v}'\cdot\overline{(\mathbf{v}'_{\mathbf{3}}\cdot\nabla_3)\mathbf{v}'}\rangle$$
$$-[\omega'K_e]_s + [\omega'K_e]_t - [\omega'\phi']_s + [\omega'\phi']_t + \langle(Residue)\rangle. \tag{4.1}$$

In this equation, the prime indicates the eddy component of the state variables and K_e is the eddy kinetic energy given by

$$K_e = \frac{1}{2}\mathbf{v}'\cdot\mathbf{v}', \tag{4.2}$$

where $\mathbf{v}'$ is the eddy component of the horizontal wind vector. The symbols ∇, ∇_3, $\mathbf{v_m}$ and $\mathbf{v}'_{\mathbf{3}}$ denote the horizontal nabla operator, the three dimensional nabla operator, the mean component of the horizontal wind vector and the eddy component of the three-dimensional wind vector for pressure vertical coordinate, respectively. Otherwise, the conventional notation is used for the state variables. The symbol $\langle\cdot\rangle$ indicates a vertical average in pressure coordinate system, and $[\cdot]$ indicates a surface integral across the surface (s) or top (t) of the model atmosphere.

The first term of the left-hand side describes the horizontal eddy kinetic energy transport, the second term is the geopotential flux convergence, the third term is the baroclinic energy conversion, and the fourth term is the barotropic energy conversion. Terms five and six describe the vertical eddy kinetic energy transport through the bottom and the top surfaces, while terms seven and eight represent the transport of eddy potential energy through the same surfaces. Finally, the last term is the residue term that represents the bulk effect of the errors of the numerical calculations and all processes unaccounted for by the other terms. The most important such process is

dissipation, which usually makes the residue term negative.

4.3 Time-mean properties of the energy conversion processes

Figure A.2 shows the time-mean eddy kinetic energy (top panel) and the time-mean of the three terms of the eddy kinetic energy that dominate the changes in the time-mean eddy kinetic energy (bottom three panels). These terms represent baroclinic energy conversion (second panel from top), barotropic energy conversion (third term from top) and the horizontal transport of the eddy kinetic energy (bottom panel).

The largest local maxima of the eddy kinetic energy are located in the eastern sector of the Pacific storm track. These maxima are due to the local generation of kinetic energy by baroclinic energy conversion and to the transport of eddy kinetic energy generated upstream by intense baroclinic energy conversion over the northwest Pacific. While barotropic energy conversion is a major sink of the eddy kinetic energy in the exit regions of the storm tracks, it is a source of the eddy kinetic energy over North America and Western Europe. We will return to this figure later when we discuss the spatial changes in the local diagnostics of ensemble performance.

5. COMPARISON OF DIAGNOSTICS FOR ALL ENSEMBLES

Figure A.3 shows the evolution of TV, TVS and VS with increasing forecast time for the ensembles included in TIGGE. We first focus our attention on studying the relationship between the evolution of VS and the evolution of TV. Then, we examine the relationship between the evolution of VS and and the evolution of TVS.

5.1 Comparison of VS and TV

A common feature of the behavior of the different ensembles at analysis time is that VS tends to be much smaller than TV. That is, the ensembles have a tendency to underestimate the analysis uncertainty. This problem is the worst for the JMA ensemble, which grossly underestimates the analysis uncertainty. The only ensemble that is practically unaffected by the problem is the CMC ensemble, for which VS is only slightly smaller than TV. Because the match between VS and TV for this ensemble is essentially perfect at all other forecast times, the slight difference at analysis time is likely to be due to the uncertainty in the proxy for the true state rather than to the underestimation of the magnitude of the initial uncertainty.

For most ensembles, VS quickly, in about 48-72 forecast hours, asymptotes to TV. The rapid recovery of the ensemble variance is particularly notable for the JMA ensemble. The unique short term behavior of this ensemble can be explained by the fact that this is the only ensemble in TIGGE that uses only right singular vectors as initial condition perturbations. Because the right singular vectors grow extremely rapidly during the optimization period, which is 48 h for the JMA ensemble, the magnitude of the analysis perturbations must be small to avoid over-shooting TV at 48 h forecast time. The ensemble that shows a somewhat similar behavior, but with a much less severe underestimation of the analysis uncertainty, is the ECMWF

ensemble. The similarity is not by accident; some of the initial condition perturbations in the ECMWF ensemble are right singular vectors. The underestimation of the uncertainty in the ECMWF ensemble is much less severe, because it mixes the right singular vectors with perturbations produced by an ensemble of data assimilations. The latter perturbations grow much slower than the right singular vectors, but their initial magnitude is larger, leading to an overall larger magnitude of the analysis perturbations.

The ensemble for which the gap between VS and TV remains relatively large at all forecast times is the CMA ensemble. This behavior is most likely due to the feature of the CMA ensemble that it is one of only two ensembles in TIGGE that does not use any "parameterization" scheme to increase the magnitude of the evolving forecast perturbations. The only other TIGGE member that does not "parameterize" the effects of model uncertainty is the KMA ensemble, but for that ensemble the gap between VS and TV is smaller than for the CMA ensemble at initial time, which seems to help at the longer forecast times as well.

5.2 Comparison of TVS and TV

According to their definitions, the diagnostics TVS and TV always satisfy the relationship $TVS \leq TV$, with smaller differences between the two indicating a better performance of the ensemble in capturing the forecast uncertainty. Similar to the behavior observed for VS and TV, the performance of the ensembles with respect to the TVS/TV relationship is the poorest at analysis time and gradually improves with forecast time. While this time there is no ensemble that would be an exception from the general trend, there are noticeable differences in the speed of the recovery of the different ensembles with forecast time.

For most ensembles, TVS saturates for long forecast times at a level that is lower

than the saturation level of TV. This is an indication of model errors for which the ensemble cannot account for. The only ensemble unaffected by this problem is the ECMWF ensemble, while the ensemble most affected is the CMA ensemble. The range of differences between the behavior of the different ensembles in this respect must be due to a combination of the differences between the severity of the model errors and the effectiveness of the ensemble generation techniques in accounting for the effects of model errors on the forecast errors.

Overall, the behavior of the operational ensembles confirms the finding of S10-11 that an operationally attainable small ensemble can efficiently span the local linear space of forecast uncertainty beyond a forecast time of about 48-72 h.

5.3 Comparison of VS and TVS

A comparison of VS and TVS can reveal whether or not an ensemble can correctly predict the total magnitude of the uncertain features it has correctly captured. The ensemble that does a particularly good job in this respect is the ECMWF, for which VS closely matches TVS at all forecast times. With the exception of the JMA ensemble, VS overestimates TVS in all other ensembles at the analysis and the short forecast times. In other words, the ensembles compensate for part of the loss of the magnitude due to not capturing some of the uncertain analysis features by over-inflating the magnitude of the uncertain analysis features that they capture correctly. For most ensembles, this strategy pays off at later forecast times in the form of a good match between VS and TVS (NCEP, UKMO and CMA) or between VS and TV (CMC and KMA). Particularly notable is the behavior of the CMC ensemble, in which the good match between VS and TV at all times is achieved by an over-inflation of the well-captured features of analysis uncertainty.

The fact that for the JMA ensemble VS remains smaller than TVS at all forecast

times suggests that the magnitude of the initial perturbations could be increased somewhat without negative effects on the performance of the ensemble.

5.4 Sensitivity of the results to the choice of the proxy for the true state

To test the robustness of our diagnostic results to the choice of the proxy for the true state, we recomputed some of the diagnostics by using analyses from a different center for the definition of $\mathbf{x}^t$. Figure A.4 shows the diagnostics for the UKMO and the CMC ensembles using ECMWF or NCEP analyses as proxies for the true state. While the results slightly change quantitatively,[1] the choice of the proxy has no effect on our qualitative observations about the the relationships between the evolution of the diagnostics.

[1]For instance, the slight underestimation of TV by VS at analysis time for the CMC ensemble is even smaller when the NCEP analyses are used as proxy for the true states.

6. COMPARISON OF DIAGNOSTICS FOR SELECTED ENSEMBLES

The diagnostic hitherto shown were computed by estimating the expected value in the definition of the diagnostics with averages over forecasts that were started at different times and over locations in the NH midlatitudes. In this section, we show maps of the ratios VS/TV and TVS/TV, estimating the expected value by averaging only over the forecasts started at different times. To save space, results are shown for only three ensembles that exhibited notable differences in their qualitative behavior in Fig. A.3. These three ensembles are those of ECMWF (Fig. A.5), NCEP (Fig. A.7) and CMC (Fig. A.9). The figures of this section has a uniform format: the panels from top to bottom show the ratio at analysis time, and 72 h, 120 h and 360 h forecast times, respectively.

6.1 ECMWF

As expected based on the results shown earlier, the ECMWF ensemble underestimates the analysis uncertainty almost everywhere (Fig. A.5). The estimate of the magnitude of the uncertainty rapidly recovers in the Atlantic storm track. On the equatorial side of the exit region of the Atlantic storm track, where zonal anomaly in the mean flow has a positive local maximum, the ensemble starts overestimating the uncertainty as early as 72 h forecast time. This region of overestimation of the uncertainty further expands at later time and at forecast time 360 h, it also covers most of the Atlantic storm track.

The recovery of the ensemble variance in the region of the Pacific storm track is slower than in the region of the Atlantic storm track. By 360 h forecast time, a local region of overestimation of the forecast uncertainty also develops on the equatorial side of the exit region of the storm track, where the zonal anomaly in the time-mean

flow has a local maximum.

It is unclear what causes the massive overestimation of the forecast uncertainty on the equatorial side of the jet region. Difficulties with representing the uncertainty in the negative barortropic energy conversion process(kinetic energy transfer from the eddies to the mean flow) are unlikely to be the cause, because the main regions of negative barotropic energy conversion are north of the problematic regions, where the ratio is about one.

The maps of the ratio TVS/TV (Fig. A.6) show a general trend that is similar to that we observed for the ratio TVS/TV: the ensemble variance recovers first in the Atlantic and Pacific storm track regions. By 360 h forecast time, the ratio is nearly one at almost all locations, which shows that at that time, the ensemble can capture almost all uncertain forecast features. This result indicates that the ECMWF implementation of the SKEB and SPPT schemes are highly efficient in representing the effects of model errors and uncertainty.

6.2 NCEP

The performance of the NCEP ensemble is clearly not as good as that of the ECMWF ensemble. It underestimates the forecast uncertainty at almost all locations and forecast times (Fig. A.7). Neither the ratio VS/TV, nor the ratio TVS/TV (Fig. A.8) reaches a value of nearly one in the storm track regions by the 120 h forecast time. At forecast time 360 h, VS/TVS reaches a value of about one only over the Atlantic storm track and the surrounding region that includes a large eastern part of North America. At the same time, TVS/TV reaches a value of about one only in the two storm track regions.

6.3 CMC

Because the ratio VS/TV (Fig. A.9) shows a unique behavior, while the ratio TVS/TV (Figure A.10) behaves very similarly to that for the NCEP ensemble, we discuss only the former diagnostics in detail. At analysis time, Fig. A.10 shows a mixture of regions of overestimation and regions of underestimation of the uncertainty. This is not an unexpected result considering that the CMC ensemble is the only one that does not underestimate the analysis uncertainty massively. Similar to the other two ensembles, the prediction of the forecast uncertainty is more accurate over the Atlantic storm track than over the Pacific storm track, and over the eastern half of North America than over the western half of North America. In contrast to the behavior of the ECMWF ensemble, there are no large regions of massive underestimation of the forecast uncertainty in the midlatitudes.

7. CONCLUSIONS

The most important findings of our study are the following:

- The finding of S10-11, that a small ensemble of forecasts is highly efficient, in a linear sense, in capturing the local space of forecast uncertainty for forecast times longer than about 72 h, are also valid for the operational global ensemble forecast system.

- The performance of the operational ensembles systems in capturing the magnitude of the forecast uncertainty is much better than what was reported by S10-11 for their research ensemble system.

- There are large difference between the performance of the different operational systems. In capturing the uncertain forecast features, the ECMWF ensemble outperforms the other ensembles. In terms of the prediction of the magnitude of the uncertainty, however, the ECMWF ensemble has the tendency to overestimate the forecast uncertainty in large regions of the midlatitudes at the longer forecast times. While the CMC ensemble provides the best prediction of the overall magnitude of the forecast uncertainty, it does that at the price of over-inflating some of the well-captured features of forecast uncertainty.

REFERENCES

Berner, J., S. Y. Ha, J. P. Hacker, A. Fournier, and C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: stochastic versus multi physics representations. *Mon. Wea. Rev.*, **139**, 1972–1995.

Berner, J., G. J. Shutts, M. Leutbecher, and T. N. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, **66**, 603–626.

Bishop, C., B. Etherton, and S. Majumdar, 2001: Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.

Bowler, N. and K. Mylne, 2009: Ensemble transform kalman filter perturbations for a regional ensemble prediction system. *QJRMS*, **135**, 757–766.

Bowler, N. E., A. Arribas, S. E. Beare, K. R. Mylne, and G. J. Shutts, 2009: The local ETKF and SKEB: Uparades to the MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **135**, 767–776.

Buizza, R., M. Leutbecher, and L. Isaksen, 2008: Potential use of an ensemble of analyses in the ecmwf ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **134**, 2051–2066.

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908.

Buizza, R., J. Tribbia, F. Molteni, and T. Palmer, 1993: Computation of optimal unstable structures for a numerical weather prediction model. *Tellus*, **45**, 388–407.

Chang, E., 2000: Wave packets and life cycles of troughs in the upper troposphere.

part ii: Seasonal and hemispheric variations. *Mon. Wea. Rev.*, **128**, 25 – 50.

Charron, M., G. Pellerin, L. Spacek, P. L. Houtekamer, N. Gagnon, H. L. Mitchell, and L. Michelin, 2010: Toward random sampling of model error in the Canadian ensemble prediction system. *Mon. Wea. Rev.*, **138**, 1877–1901.

Houtekamer, P. L., 2002: The use of multiple parameterizations in ensembles. *ECMWF Workshop on Model Uncertainty, 20–24 June 2011*, Reading, UK, ECMWF, 163–174.

Lorenz, E. N., 1969: Predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307.

Molteni, F. and T. Palmer, 1993: Predictability and finite-time instability of the northern winter circulation. *Quart. J. Roy. Meteor. Soc.*, **119**, 269–298.

Mureau, R., F. Molteni, and T. Palmer, 1993: Ensemble prediction using dynamically conditioned perturbations. *Quart. J. Roy. Meteor. Soc.*, **119**, 299–323.

Orlanski, I. and E. Chang, 1993: Ageostrophic geopotential fluxes and upstream development of baroclinic waves. *J. Atmos. Sci.*, **50**, 212–225.

Orlanski, I. and J. Katzfey, 1991: The life cycle of a cyclone wave in the southern hemisphere. part i: Eddy energy budget. *J. Atmos. Sci.*, **48**, 1972–1998.

Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parameterization and model uncertainty. ECMWF Technical Memorandum, number = 598, pages = 42.

Rotunno, R. and C. Snyder, 2008: A generalization of lorenz's model for the predictability of flows with many scales of motion. *J. Atmos. Sci.*, **65**, 1063–1076.

Satterfield, E. and I. Szunyogh, 2010: Predictability of the performance of an ensemble forecast system: Ppredictability of the space of uncertainties. *Mon. Wea. Rev.*, **138**, 962–981.

Satterfield, E. and I. Szunyogh, 2011: Assessing the performance of an ensemble

forecast system in predicting the magnitude and the spectrum of analysis and forecast uncertainties. *Mon. Wea. Rev.*, **139**, 1207–1223.

Shutts, G. J., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102.

Shutts, G. J., 2013: Coarse graining the vorticity equation in the ECMWF integrated forecasting system: the search for kinetic energy backscatter. *J. Atmos. Sci.*, **70**, 1233–1241.

Talagrand, O., 1981: A study of the dynamics of four-dimensional data assimilation. *Tellus*, **33**, 43–60.

Tennant, W. J., G. J. Shutts, A. Arribas, and S. A. Thompson, 2011: Using a stochastic kinetic energy backscatter scheme to improve MOGREPS probabilistic forecast skill. *Mon. Wea. Rev.*, **139**, 1190–1206.

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at nmc: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.

Toth, Z. and E. Kalnay, 1997: Ensemble forecasting at ncep and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.

Tribbia, J. J. and D. P. Baumhefner, 2004: Scale interactions and atmospheric predictability: an updated perspective. *Mon. Wea. Rev.*, **132**, 703–713.

Wang, X. and C. Bishop, 2003: A comparison of breeding and ensemble transform kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158.

Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the ensemble transform (et) technique in the ncep global operational forecast system. *Tellus*, **60**, 62–79.

APPENDIX A

FIGURES

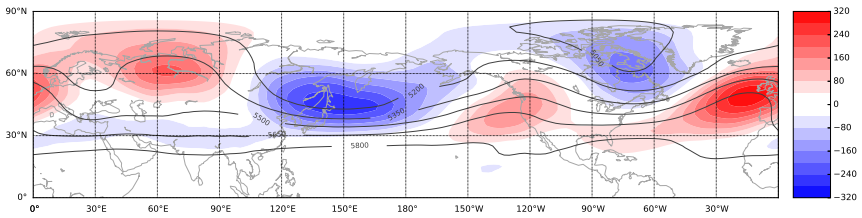

Figure A.1: Contours indicate the time-mean flow used for the definition of the eddy components of the state variables in the computation of the terms of the eddy kinetic energy equation. Color shades show zonal anomalies in the time mean flow that have an amplitude larger 40 gpm.

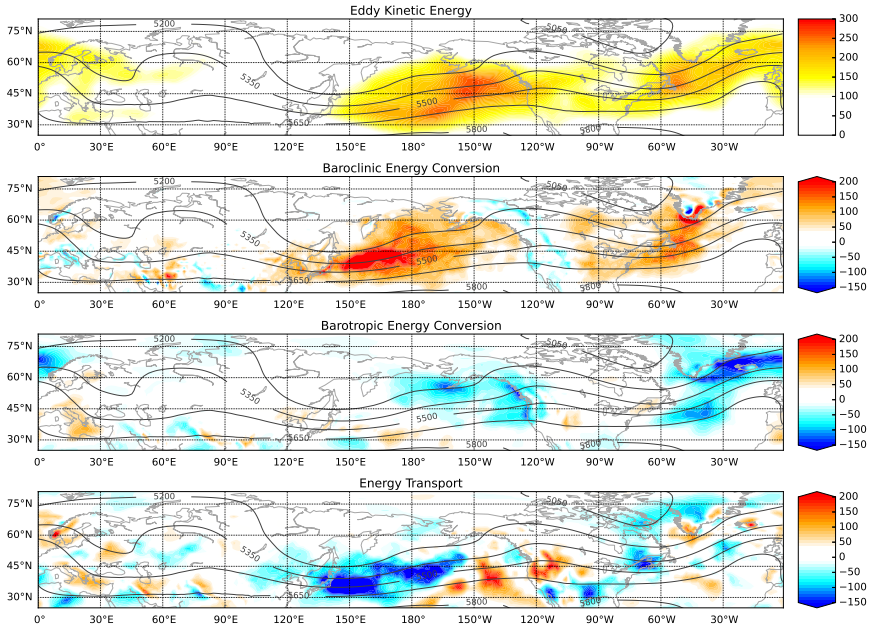

Figure A.2: The time-mean of the eddy kinetic energy conversion processes for January-February-March 2012. Shown by color shades are the time mean of the (top) eddy kinetic energy (second from top) baroclinic energy conversion (third from top) barotropic energy conversion and (bottom) horizontal transport of the eddy kinetic energy.

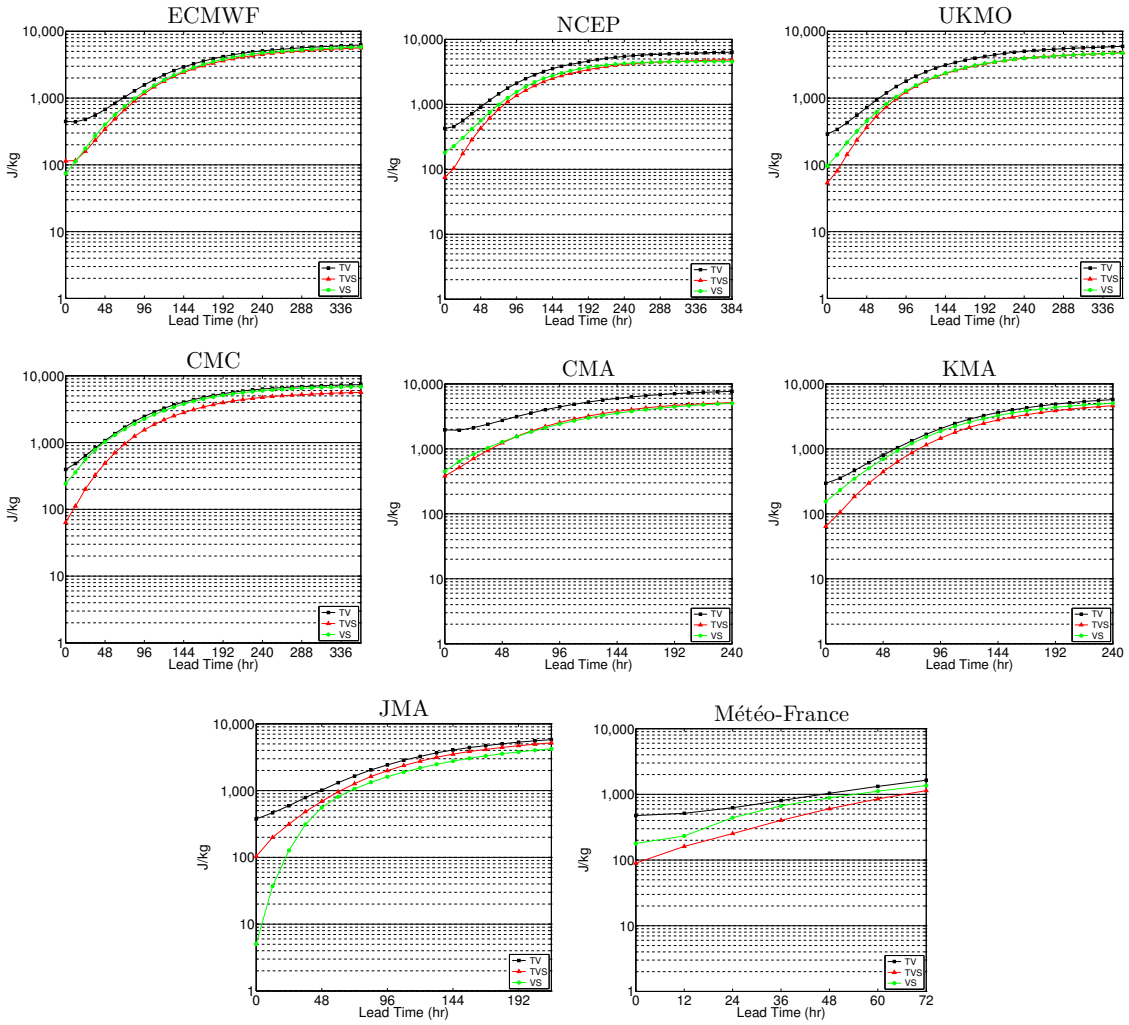

Figure A.3: V, TV, and TVS for each ensemble prediction system, averaged for the northern hemisphere extra-tropics and all forecasts initialized between January 1, 2012 and February 29, 2012.

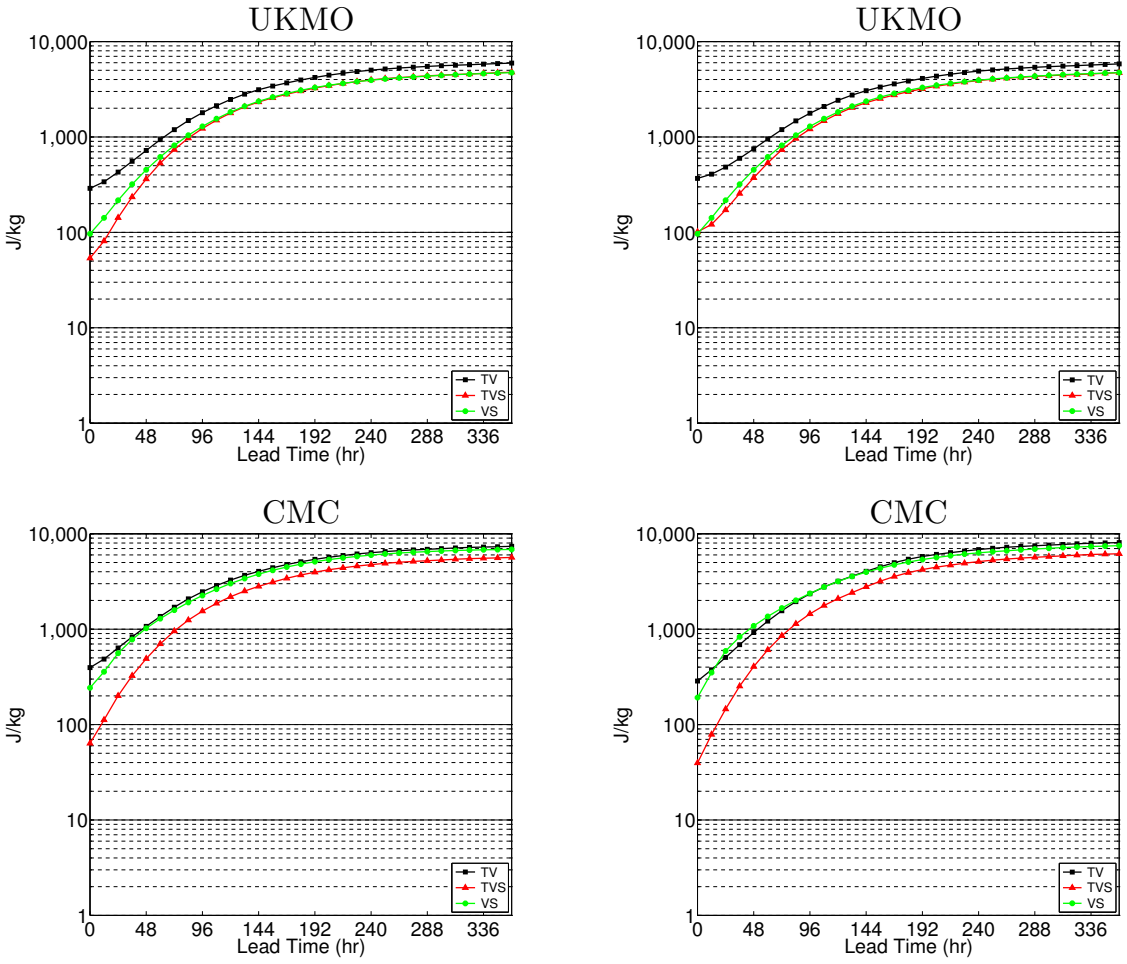

Figure A.4: Comparing the diagnostics of the UKMO and CMC ensembles using two different verifications, ECMWF verification (left) and NCEP verification (right).

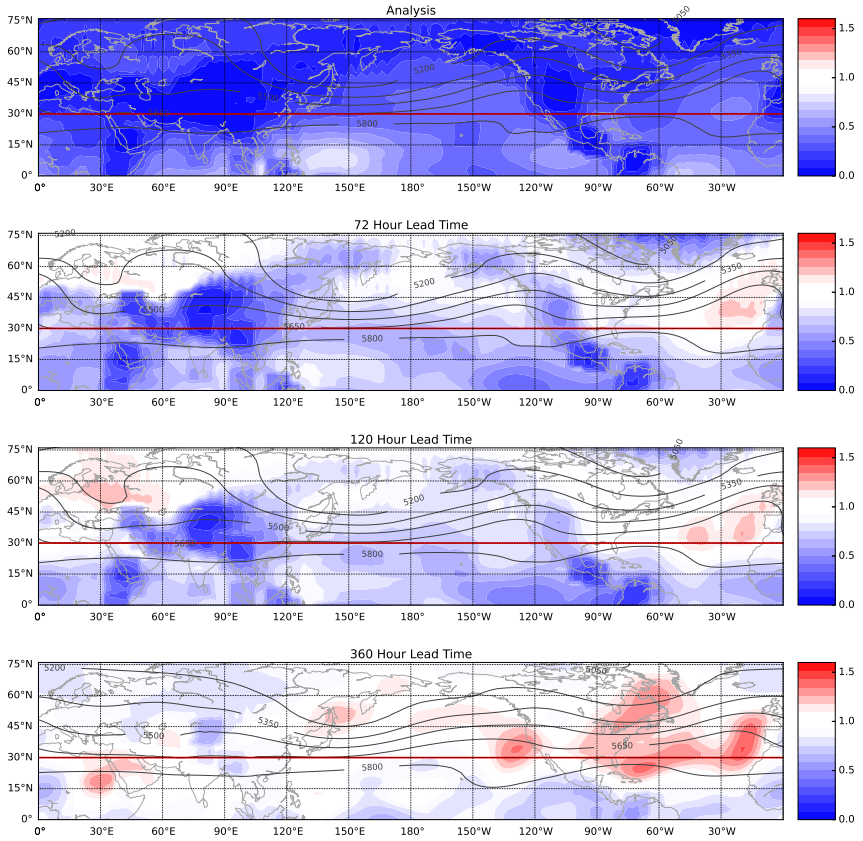

Figure A.5: Geographical distribution of ensemble performance (VS/TV) for the ECMWF ensemble system, averaged over January and February, 2012.

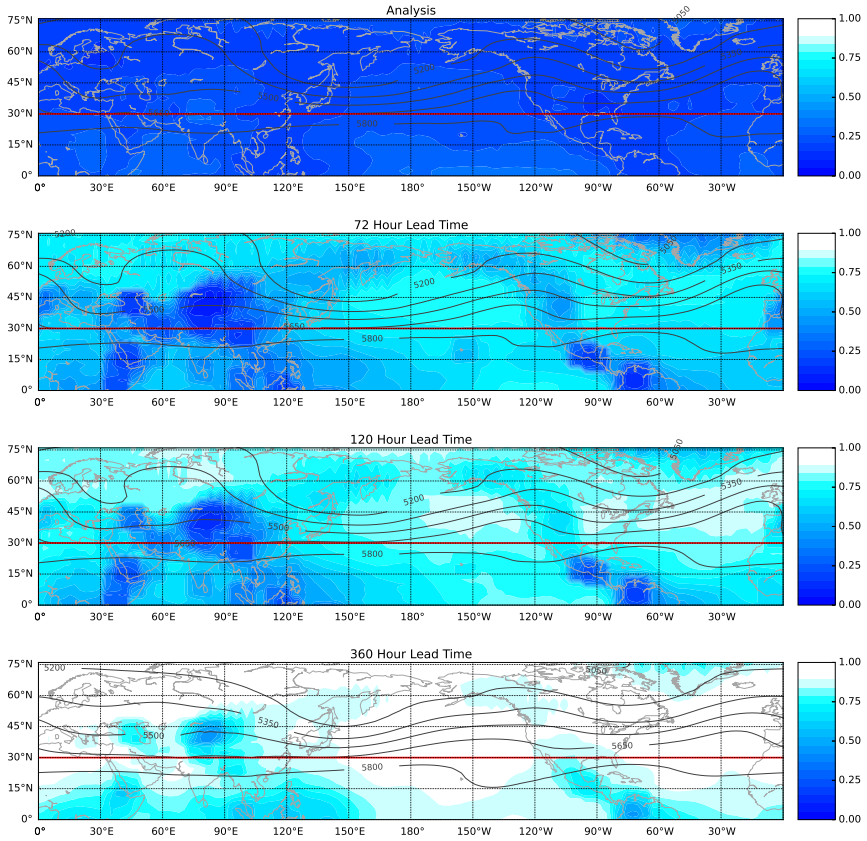

Figure A.6: Geographical distribution of TVS/TV for the ECMWF ensemble system, averaged over January and February, 2012.

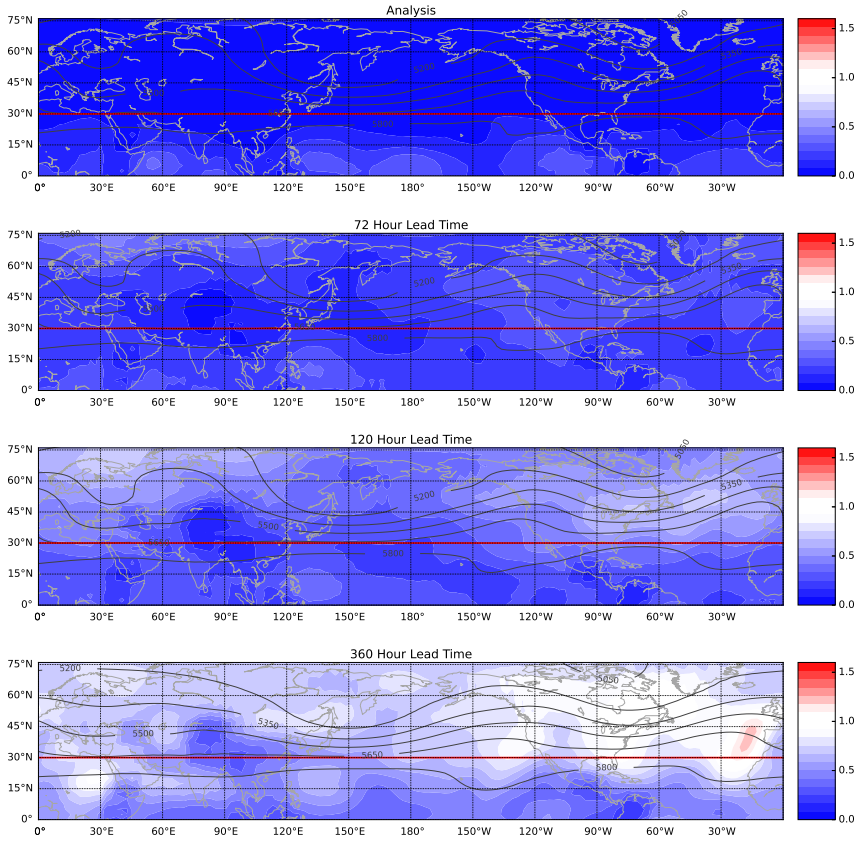

Figure A.7: Geographical distribution of ensemble performance (VS/TV) for the NCEP ensemble system, averaged over January and February, 2012.

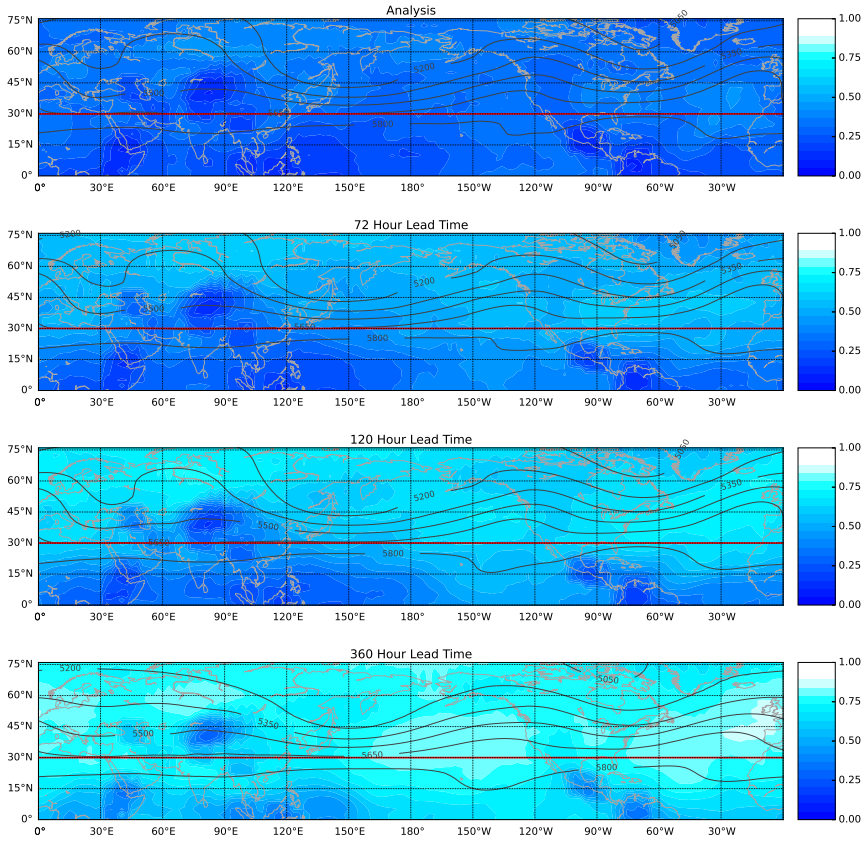

Figure A.8: Geographical distribution of TVS/TV for the NCEP ensemble system, averaged over January and February, 2012.

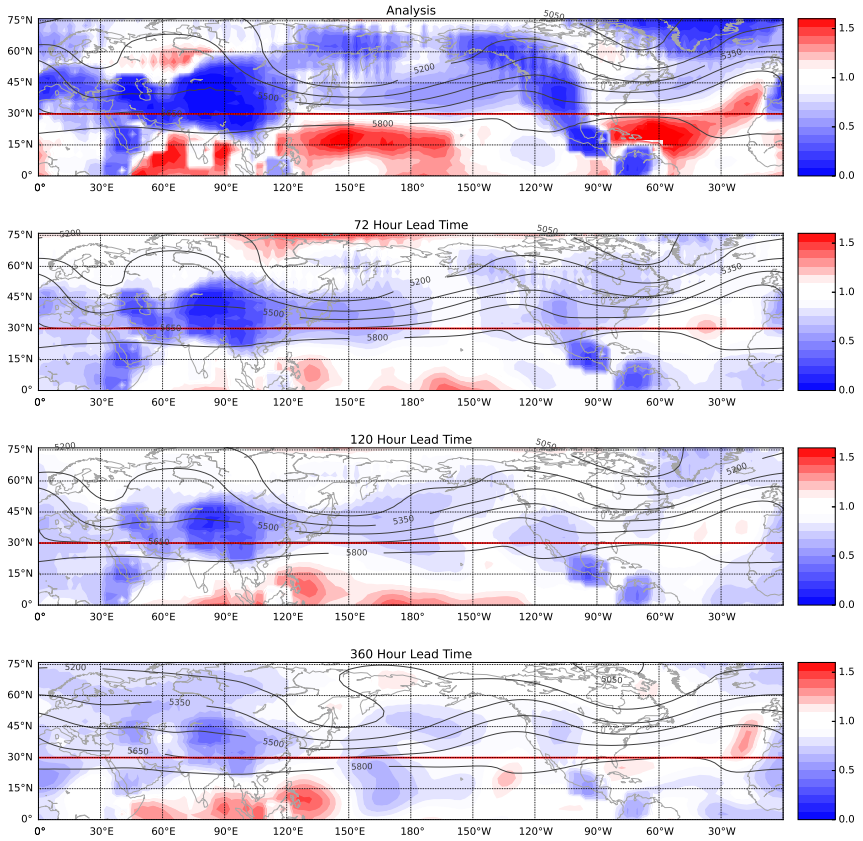

Figure A.9: Geographical distribution of ensemble performance (VS/TV) for the CMC ensemble system, averaged over January and February, 2012.

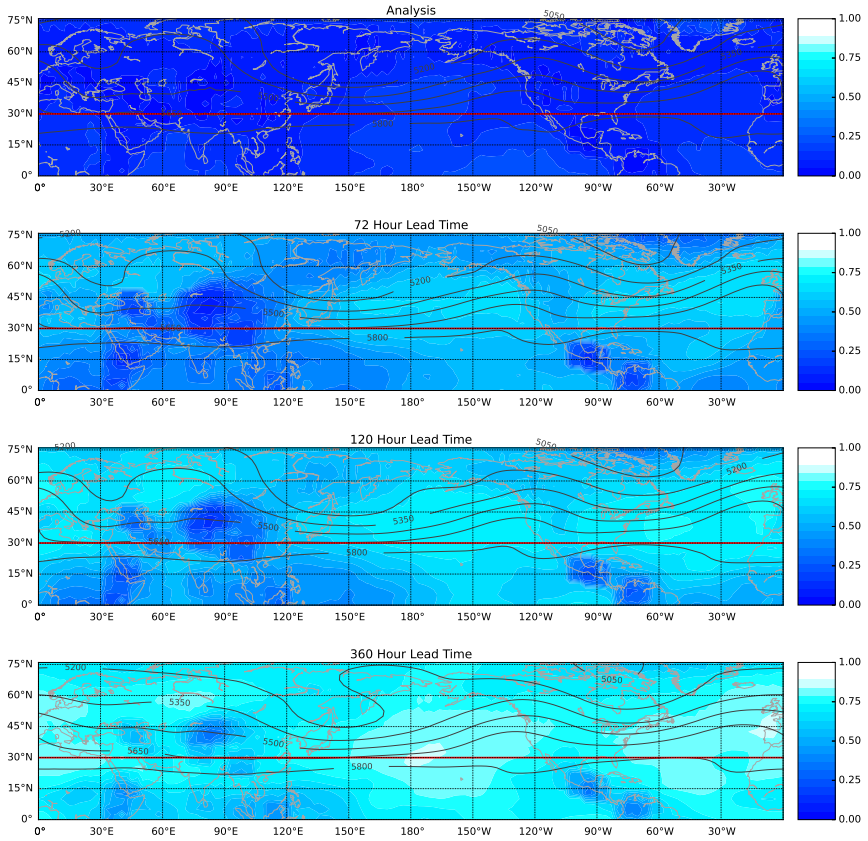

Figure A.10: Geographical distribution of TVS/TV for the CMC ensemble system, averaged over January and February, 2012.

APPENDIX B

TABLES

Table B.1: Ensemble Forecast Systems Included From TIGGE

NWP Center	Representation of Model Error and Uncertainty	Initial Perturbation Strategy
ECMWF	SKEB/SPPT	Singular Vectors & EDA
NCEP	SPPT	Ensemble Transform & Rescaling
UKMO	SKEB	ETKF
CMA	None	Bred Vectors
CMC	SKEB/SPPT	EnKF
KMA	None	Bred Vectors
JMA	SPPT	Singular Vectors
Météo-France	Multi-Physics	Singular Vectors & Evolved Singular Vectors